Emergency appendectomy is the most common emergency abdominal procedure performed in the United States.\(^1\) Although patients with acute appendicitis often present with a characteristic symptom complex and physical findings, atypical presentations are common. Missed or delayed diagnosis can lead to increased rates of perforation, morbidity, and unnecessary appendectomy. In 1997, 15.3% of appendectomies in the United States were judged to be unnecessary, resulting in $741.5 million of potentially avoided hospital charges.\(^2\) Patients with negative appendectomies have a higher mortality rate than those with acute appendicitis, possibly due to the delay in delineation of the actual cause of presenting symptoms.\(^2\)

History and physical examination are at the foundation of the diagnosis of acute appendicitis. However, due to the imprecision of these findings, adjunctive studies are usually obtained. The prevailing perception is that the typical patient with appendicitis will have a modest elevation of temperature. Most physicians evaluating patients with presumed appendicitis obtain a complete blood cell count, with the expectation that elevation of the total white blood
cell (WBC) count supports the diagnosis of appendicitis, and a normal or low count tends to mitigate the diagnosis. The most recent editions of several leading medical texts recommend this test as part of the evaluation of patients with possible appendicitis, despite several studies that acknowledge the limitations of this test.

Most studies investigating the utility of total WBC count have been retrospective and have comprised patients who were operated on or admitted to the hospital for presumed acute appendicitis. Investigating the utility of total WBC count on this relatively narrow spectrum of patients can introduce bias and does not reflect the real-world application of the test in a more diverse patient population in a clinical setting such as the emergency department (ED).

The objective of this study was to assess the discriminatory value of total WBC count and temperature in a consecutive group of patients presenting to an ED with signs and symptoms suggestive of acute appendicitis.

**METHODS**

**Study Design.** This was a prospective study performed in the University of California at San Diego Medical Center ED, with an annual census of 35,000 patients. The institutional review board approved this study. Informed consent was obtained from all study participants.

**Study Setting and Population.** Patients were enrolled in consecutive fashion between April 1998 and March 2000. The study was designed to assess the impact of a management guideline utilizing helical computed tomography (CT) as an imaging adjunct in patients for whom the diagnosis of appendicitis was considered. This investigation represents a planned subanalysis of the total WBC count and temperature measurements obtained as part of the routine data collected on enrolled patients.

**Study Protocol.** The study protocol called for the inclusion of all patients presenting with signs and symptoms that led the managing attending physician in the ED to conclude that acute appendicitis was the primary diagnostic consideration. There were no age exclusions; however, due to the possible use of x-ray as part of the study protocol, all pregnant patients were excluded. Patients were enrolled in a consecutive manner after informed consent was obtained. In the circumstance of a minor, consent was obtained from an appropriate guardian. All laboratory tests were obtained at the discretion of the managing physicians. Admitted patients were followed up to final surgical or clinical outcome. All patients discharged from the ED were followed up by telephone interview two weeks after the index visit.

**Data Analysis.** Analysis of the data on total WBC count and temperature included determination of sensitivities, specificities, positive predictive values (PPVs), negative predictive values (NPVs), positive likelihood ratios (+LRs), and negative likelihood ratios (−LRs). All results were reported with 95% confidence intervals (95% CIs). Receiver-operating characteristic (ROC) curves were plotted for these variables along with calculation of the area under the curve (AUC) with 95% CI. Frequency histograms were developed and plotted for total WBC count and temperature values and the presence or absence of appendicitis. These methods of data analysis allowed for the assessment of a range of total WBC count and temperature values beyond the measures selected on the a priori basis mentioned above. All statistical analysis was performed using StatsDirect version 1.9.8 (StatsDirect Software, Cheshire, England).

**RESULTS**

A total of 308 patients were enrolled between April 1998 and March 2000; however, 15 were lost to follow-up and excluded from the study, yielding a final study population of 293 patients. There were 130 male patients and 163 female patients. The mean age of the patients was 30.8 years (range, 7–75 years). Ninety-two of the 293 patients had appendicitis, for a prevalence of 31%. Perforated appendix was identified in 21% of patients with appendicitis. Helical CT was performed in 248 patients, and 60 patients were admitted to the surgical service directly without CT imaging of the abdomen. The results of helical CT scans have been previously reported.

Total WBC count was measured in 274 patients. Table 1 shows a standard $2 \times 2$ table for total WBC count in patients with and without appendicitis. Total WBC count $>10,000$ cells/mm$^3$ yielded a sensitivity
of 76% (95% CI = 65% to 84%) and a specificity of 52% (95% CI = 45% to 60%). The PPV was 42% (95% CI = 35% to 51%), and the NPV was 82% (95% CI = 74% to 89%). The \( \text{LR} \) was 1.59 (95% CI = 1.31 to 1.93), and the \( \text{LR} \) was 0.46 (95% CI = 0.31 to 0.67).

Table 2 shows the LRs for different interval values of total WBC count.

Figure 1 shows an ROC curve for total WBC count. The AUC for the ROC curve was 0.72 (95% CI = 0.65 to 0.79). If one chose to maximize both sensitivity and specificity, an optimum cutoff for an abnormal total WBC count would be set at 11,900 cells/mm\(^3\), which would yield a sensitivity of 69% and a specificity of 72% (Figure 1).

Temperature was recorded for all 293 patients. Table 3 shows a standard 2 x 2 table for temperature in patients with and without appendicitis. The sensitivity of an elevated temperature for appendicitis was 47% (95% CI = 36% to 57%), and the specificity was 64% (95% CI = 57% to 71%). The PPV was 37% (95% CI = 29% to 46%), and the NPV was 72% (95% CI = 65% to 79%). The \( \text{LR} \) was 1.3 (95% CI = 0.97 to 1.72), and the \( \text{LR} \) was 0.82 (95% CI = 0.65 to 1.01). Table 4 shows LRs for different interval values of temperature. This analysis shows that patients with a temperature less than 98\(^\circ\)F were slightly less likely to have appendicitis, with a \( \text{LR} \) of 0.64 (95% CI = 0.41 to 0.95). The presence of a fever had minimal correlation with a diagnosis of appendicitis (Table 4). Figure 2 shows an ROC curve analysis for temperature, with an AUC of 0.59 (95% CI = 0.52 to 0.66).

**DISCUSSION**

Measurement of total WBC count and temperature is usually considered a routine part of the workup for acute appendicitis. Many emergency physicians have observed a surgical consultant discounting the diagnosis of acute appendicitis in a particular patient because of a normal total WBC count or temperature. The data in this study indicate that neither total WBC count nor temperature is a useful indicator of the presence or absence of acute appendicitis in an at-risk population.

In this study, we analyzed the diagnostic value of total WBC count and temperature in ED patients with signs and symptoms suggestive of acute appendicitis. For a cutoff value of 10,000 cells/mm\(^3\), the sensitivity of the total WBC count is modest (76%) and the specificity is poor (47%). Additionally, the ROC curve analysis suggests that there is no value of total WBC count that is sensitive and specific enough to be clinically useful. An ideal test has an AUC of 1, while a perfectly random test has an AUC of 0.5. Generally, a “good” test has an AUC >0.8 and an “excellent” test has an AUC >0.9.

Previous studies assessing the relationship between total WBC count and appendicitis have their findings reported in a variety of ways, including comparing mean values for total WBC count in patients with and without appendicitis, and variously using p-values and at times sensitivity, specificity, PPV, and NPV.

---

**TABLE 1. Total WBC Count and Appendicitis**

<table>
<thead>
<tr>
<th>Total WBC (1,000 cells/mm(^3))</th>
<th>Not Appendicitis</th>
<th>Appendicitis</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤10,000</td>
<td>98 (52%)</td>
<td>21 (24%)</td>
<td>199</td>
</tr>
<tr>
<td>&gt;10,000</td>
<td>89 (48%)</td>
<td>66 (76%)</td>
<td>155</td>
</tr>
<tr>
<td>Totals</td>
<td>187</td>
<td>87</td>
<td>274</td>
</tr>
</tbody>
</table>

A standard 2 x 2 table for patients with and without appendicitis. For the purposes of this analysis, a WBC count >10,000 cells/mm\(^3\) was considered abnormal.

**TABLE 2. Likelihood Ratios for Different Interval Values of WBC**

<table>
<thead>
<tr>
<th>WBC Count (1,000 cells/mm(^3))</th>
<th>Not Appendicitis</th>
<th>Appendicitis</th>
<th>Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤8</td>
<td>57</td>
<td>11</td>
<td>0.41 (0.23, 0.72)</td>
</tr>
<tr>
<td>8 to &lt;10</td>
<td>41</td>
<td>10</td>
<td>0.52 (0.27, 0.96)</td>
</tr>
<tr>
<td>10 to &lt;12</td>
<td>38</td>
<td>6</td>
<td>0.34 (0.15, 0.73)</td>
</tr>
<tr>
<td>12 to &lt;15</td>
<td>33</td>
<td>26</td>
<td>2.70 (1.79, 4.20)</td>
</tr>
<tr>
<td>≥15</td>
<td>25</td>
<td>28</td>
<td>2.38 (1.48, 3.81)</td>
</tr>
</tbody>
</table>

**TABLE 3. Temperature and Appendicitis**

<table>
<thead>
<tr>
<th>Temperature (°F)</th>
<th>Not Appendicitis</th>
<th>Appendicitis</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;99.0</td>
<td>129 (64%)</td>
<td>49 (53%)</td>
<td>178</td>
</tr>
<tr>
<td>≥99.0</td>
<td>72 (36%)</td>
<td>43 (47%)</td>
<td>115</td>
</tr>
<tr>
<td>Totals</td>
<td>201</td>
<td>92</td>
<td>193</td>
</tr>
</tbody>
</table>

A standard 2 x 2 table for patients with and without appendicitis. For the purposes of this analysis, a temperature ≥99°F was considered abnormal.
These studies can be difficult to interpret, because both PPV and NPV depend on the prevalence of the disease. Moreover, sensitivity and specificity alone do not allow clinicians to directly apply the results of diagnostic tests to individual patients. Additionally, a finding of statistical significance (p < 0.05) may not always translate to clinical significance.

The LR is defined as the true-positive rate over the false-positive rate, or sensitivity/(1−specificity). It allows the clinician to assess the likelihood that a patient with a given test result (i.e., elevated total WBC count) has that disease. Additionally, unlike PPV and NPV, the LR is independent of disease prevalence, which allows for some comparison of a diagnostic test's performances among different populations. Generally, a clinically useful diagnostic test has an LR > 10 or < 0.1. In that light, our finding of a +LR of 1.59 for a total WBC count > 10,000 cells/mm³ and a −LR of 0.46 for a total WBC count ≤ 10,000 cells/mm³, while statistically significant, is actually quite modest and unlikely to be clinically useful.

Calculating a single LR from a continuous variable such as total WBC count causes the loss of significant information and may actually overestimate the utility of a diagnostic test. Such an analysis forces a continuous variable, such as the total WBC count, to become binary. One solution to this problem is to divide the variable into intervals and calculate an LR for each interval (Table 2). When we do this for total WBC count, we find that the value of the LR increases slightly, with a highest value of 2.70. However, even this LR remains relatively low based on the definition previously offered and is unlikely to be significantly more useful than the value of 10,000 cells/mm³ used in the empirical analysis.

The limited utility of a modest LR is best demonstrated by an example. Consider a patient in the ED presenting with right lower quadrant pain, for whom the treating physician has a suspicion of appendicitis. This hypothetical patient does not have peritoneal signs but does have significant right lower quadrant tenderness to palpation. There are, however, some atypical features about the presentation because the patient has had pain for only a few hours and feels hungry. In this setting, we will assume the treating physician assigns a pretest probability of disease at approximately 25%. A total WBC count is ordered and returns with a value of 13,300 cells/mm³. This corresponds to a +LR of 2.69 (Table 2). Applying this LR to the Fagan nomogram yields a posttest probability of 47%. This level of total WBC count increases the probability of appendicitis in this patient from 25% to 47%, but this is not sufficiently different enough from the pretest probability to compel a surgeon who is hoping to avoid unnecessary surgery to operate on the patient. If our patient had a total WBC count of 7,000 cells/mm³, the −LR would be 0.41 (Table 3), which yields a posttest probability of 15%, significantly lower than the pretest value but not sufficiently low enough to safely discard the diagnosis of appendicitis or discharge the patient. The nature of LRs is such that they are most useful for patients with intermediate pretest probability, so changing the pretest probability of our example scenario would not substantially alter the impact of the findings.

While calculating the LR for different interval values of total WBC count yields more information than simply choosing a single cutoff value, potentially significant information can still be lost as one transforms a continuous variable into a set of a few intervals. One way around this problem is to plot frequency histograms of the patients with and without appendicitis and their corresponding total WBC count values (Figure 3). In this analysis, the ratio of the height of the histogram of those with appendicitis to the height of those without appendicitis is the LR for that particular value of total WBC count (i.e., true positives over false positives). One could then theoretically find an LR for any value of total WBC count and appendicitis. Additional precision can be gained by fitting curves to the data.

The main difficulty with this type of analysis is that the number of patients may be small for any given

### Table 4. Likelihood Ratios for Different Interval Values of Temperature

<table>
<thead>
<tr>
<th>Temperature (°F)</th>
<th>Not Appendicitis</th>
<th>Appendicitis</th>
<th>Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;98.0</td>
<td>70</td>
<td>21</td>
<td>0.64 (0.41, 0.95)</td>
</tr>
<tr>
<td>98.0 to &lt;99.0</td>
<td>59</td>
<td>28</td>
<td>1.00 (0.68, 1.45)</td>
</tr>
<tr>
<td>99.0 to &lt;100.0</td>
<td>45</td>
<td>18</td>
<td>0.85 (0.52, 1.36)</td>
</tr>
<tr>
<td>100.0 to &lt;101.0</td>
<td>13</td>
<td>12</td>
<td>1.95 (0.94, 4.03)</td>
</tr>
<tr>
<td>101.0 to &lt;102.0</td>
<td>8</td>
<td>9</td>
<td>2.38 (0.97, 5.79)</td>
</tr>
<tr>
<td>≥102.0</td>
<td>4</td>
<td>6</td>
<td>3.18 (0.98, 10.25)</td>
</tr>
</tbody>
</table>

Figure 2. Receiver-operating characteristic curve for temperature and appendicitis. The area under the curve (AUC) is 0.59 (95% CI = 0.52 to 0.66). An ideal test has an AUC of 1.00, while a perfectly random test has an AUC of 0.5.
value of total WBC count, resulting in large CIs. The calculation of CIs for the quotient of distribution functions is a complex task not readily accessible to the clinician. For this reason, Tandberg et al. advocate the incorporation of such analysis into the routine process of computerized laboratory result retrieval.\(^2\) The computer could then calculate and display disease-specific likelihoods for each possible laboratory value. In an ideal test, the populations of those with and without the disease have no overlap (Figure 4), while in the real world a good diagnostic test demonstrates minimal overlap that generally translates to an LR > 10 or < 0.1 at an assigned cutoff value. In the circumstance of this study, even such a sophisticated analysis does not allow the total WBC count to be clinically useful in the diagnosis of acute appendicitis. Figure 3 demonstrates that the distributions of the total WBC count among those with and without appendicitis exhibit a very high degree of overlap.

Other investigators have constructed ROC curves for total WBC count and appendicitis with similar results. Korner et al. found an AUC of 0.69 (95% CI = 0.65 to 0.73), statistically no different from the results presented here.\(^2\) That study was performed on patients operated on for suspected appendicitis, while our data are for ED patients with signs and symptoms suggestive of appendicitis. However, ROC curve analysis is independent of prevalence, allowing for a comparison of data. Granoos et al. found an AUC of 0.730 (standard error, 0.041).\(^3\) Rodriguez-Sanjuan et al. found an AUC of 0.67 (standard error, 0.08; CI not available) for total WBC count and appendicitis in children.\(^3\) Paajanen et al. found an AUC of 0.76 (no CI reported).\(^3\) Andersson et al. found an AUC of 0.80 (standard error, 0.02, CI not available) for patients admitted to the hospital for suspected appendicitis.\(^3\) This value is significantly higher than other reported values, including our result of 0.72. The reason for this discrepancy is not clear, although it may have to do with the difference in study patients. The study population for Andersson et al. consisted of patients in Sweden admitted to the hospital for suspected appendicitis and is probably not based on the initial but rather a delayed blood count measure, as demonstrated in a report published one year later on the same population.\(^3\) This might account for the difference in AUC of the ROC curves in that study compared with others.

Using the data of Dueholm et al., Snyder and Hayden calculated LRs for defined intervals of total WBC counts.\(^3\) In general, their results are similar to those presented here; however, in that study, the total WBC count appears to discriminate slightly better than here. For example, the LR for a total WBC count between 4,000 and 7,000 cells/mm\(^3\) is 0.10 (95% CI = 0 to 0.39). However, because the numbers involved are small, the CIs are wide. Andersson et al. also calculated LRs at the same intervals presented in this study.\(^3\) As with their ROC curve data, their results suggest that total WBC count is a better discriminator than the data presented here. They found an LR of 0.16 (95% CI = 0.10 to 0.26) for total WBC count < 8,000 cells/mm\(^3\) and an LR of 7.03 (95% CI = 4.11 to 12.15) for total WBC count > 15,000 cells/mm\(^3\), significantly greater than our reported values of 0.41 (95% CI = 0.23 to 0.72) and 2.38 (95% CI = 1.48 to 3.81), respectively. The reasons for this may again be related to the nature of the patient population and the use of repeated total WBC count measures over time during observation in an inpatient setting.

One might query as to what harm total WBC count might have, even if it has little discriminatory value. While largely speculative, it is possible that an elevated total WBC count might erroneously lead...
appendicitis, it may give clues to clinicians about the presence of other infectious diseases.

CONCLUSIONS

While there is a statistically significant relationship between total WBC count and acute appendicitis in ED patients with signs and symptoms suggestive of acute appendicitis, this relationship is modest and not believed to be clinically useful. We caution against reliance on measurement of total WBC count in the evaluation of ED patients with suspected appendicitis. Body temperature is not a useful discriminator in the evaluation of ED patients with suspected appendicitis.

LIMITATIONS

Treating clinicians were not blinded as to the values of total WBC count and temperature. This lack of blinding can lead to several types of bias in associating the test result and the disease, including workup bias (the test result affects subsequent clinical workup). Workup bias can lead to underdiagnosis, thereby making a test appear more sensitive than it actually is. However, because all patients not operated on for appendicitis or admitted to the hospital were followed up by telephone 14 days after presentation, this type of bias is unlikely.

Additionally, we did not study the possible diagnostic value of the neutrophil count. Previous studies have not found the neutrophil count to be superior to total WBC count in the diagnosis of acute appendicitis. We also did not study how total WBC count might perform as part of a scoring system or combination of test results. Finally, while total WBC count alone does not appear to be clinically useful in discriminating between patients with and without appendicitis, it may give clues to clinicians about the presence of other infectious diseases.

Figure 5. Temperature distribution of patients with and without appendicitis. The y-axis shows the number of cases. There is almost complete overlap of the distributions.

**REFERENCES**